KU LEUVEN

FACULTY OF ECONOMICS AND BUSINESS

Volatility spillovers and heavy tails: a large *t*-Vector AutoRegressive approach

Barbaglia L, Croux C, Wilms I.

KBI_1716

Volatility Spillovers and Heavy Tails: A Large t-Vector AutoRegressive Approach

Luca Barbaglia[a,*], Christophe Croux[a] and Ines Wilms[a,c]

[a] *Faculty of Economics and Business, KU Leuven, Belgium*

[c] *Cornell University, Ithaca, USA*

Abstract. Volatility is a key measure of risk in financial analysis. The high volatility of one financial asset today could affect the volatility of another asset tomorrow. These lagged effects among volatilities - which we call volatility spillovers - are studied using the Vector AutoRegressive (VAR) model. We account for the possible fat-tailed distribution of the VAR model errors using a VAR model with errors following a *multivariate Student* t-distribution with unknown degrees of freedom. Moreover, we study volatility spillovers among a *large* number of assets. To this end, we use *penalized* estimation of the VAR model with t-distributed errors. We study volatility spillovers among energy, biofuel and agricultural commodities and reveal bidirectional volatility spillovers between energy and biofuel, and between energy and agricultural commodities.

Keywords: Commodities, Forecasting, Multivariate t-distribution, Vector AutoRegressive model, Volatility spillover.

JEL classification: C58, C32, Q02

*Corresponding author: luca.barbaglia@kuleuven.be

1 Introduction

Volatilities of financial asset returns are closely followed by financial market analysts and investors. As volatility is a risk measure, analysts working in portfolio management are interested in better understanding the volatility of the financial assets in their portfolio in order to minimize the exposure to future potential losses. Knowing whether the high volatility of one financial asset today leads to high volatility of another asset tomorrow can be key information to reduce this risk exposure. We focus on these lagged effects among volatilities and we refer to them as volatility spillovers.

We investigate the existence of volatility spillovers among energy, biofuel and agricultural commodities (see Serra and Zilberman, 2013, and references therein). Commodity markets are of great interest for financial analysts and investors, who uses them for risk hedging purposes or as alternative investment areas. In recent years, the high volume of financial transactions in commodity markets has lead to a detachment from simple supply-demand dynamics. This makes it harder for analysts to describe swings in volatilities simply recurring to the economic fundamentals. In addition, commodity markets are volatile by their very nature: a policy change, a natural disaster or a technological breakthrough might cause a unexpected period of high volatility.

Traditionally, high volatility in the energy markets would affect unidirectionally the volatility of agricultural products via energy intensive agricultural inputs (e.g., fuel and fertilizers). The emergence of biofuel production has changed the link between energy and agricultural commodities as certain crops represent simultaneously food, feedstock and fuel sources. A joint modeling of the volatility spillovers among energy, biofuel and agricultural commodities is therefore relevant, as it might improve traditional risk management tools and lead to a better assessment of biofuel support policies (Serra, 2011).

We follow Diebold and Yilmaz (2015) and analyze volatility spillovers using the Vector AutoRegressive (VAR) model. First, we obtain a measure of volatility (e.g. see Martens and van Dick, 2007 or McAleer and Medeiros, 2008). Then, we consider a VAR model containing the logarithmic transformed volatilities as time series to estimate the spillovers. This procedure has been used in many empirical applications, but two main concerns remain: (i) the standard estimation procedure of the VAR does not account for fat-tailed errors and thus for the frequent occurrence of extreme observations in the volatility series (e.g. Callot et al., 2017) and, (ii) the number of time series in the VAR model is limited since the number of parameters to be estimated increases quadratically with the number of time series included (Diebold and Yilmaz, 2015, p183).

Several recent papers propose methods to estimate large VAR models - that is models containing a large number of time series relative to the time series length - (e.g. Basu and Michailidis, 2015; Davis et al., 2016; Gelper et al., 2016; Barigozzi and Hallin, 2017; Derimer et al., 2017). The extension of large VAR models to fat-tailed errors has, however, not been addressed yet. It has been argued that log volatilities

are approximately normal (Diebold and Yilmaz, 2015, p18), which is supported by empirical evidence (e.g. Andersen et al., 2001) and asymptotic theory (Barndorff-Nielsen and Shephard, 2002). Nevertheless, subsequent studies emphasize that, in practice, this might not hold for all volatility measures. Considering realized variances, Corsi et al. (2008) and Hassler et al. (2016) highlight that the log volatility of the S&P 500 index based on 5min intra-day returns exhibits deviations from the Gaussian distribution. Considering realized ranges, Christensen et al. (2009) prove that the log volatility follows a mixed multivariate normal, and Caporin and Velo (2015) show that the log volatility has heavier tails than the normal. Hence, it is appropriate to specify a more general distribution for the errors of the VAR model.

We analyze volatility spillovers using a large t-VAR estimation approach. To this end, we rely on the work of Diebold and Yilmaz (2015), but we (i) use a VAR model with *t-distributed errors*. We *estimate the degrees of freedom* of the t-distribution, in contrast to prior studies who take them to be fixed (e.g. Franses and Lucas, 1998; Finegold and Drton, 2011). As such, we determine the degree of fat-tailedness of the VAR residuals in a data-driven manner. (ii) We consider *large* t-VAR models containing a large number of time series relative to the time series length. To ensure that the estimation of this large VAR model is feasible, we use a penalized estimator of the VAR model in the spirit of the Lasso (Tibshirani, 1996) and call it the *t-Lasso*. The t-Lasso attains higher estimation and forecast accuracy in presence of fat-tailed errors than other standard estimators, as confirmed by our simulation study.

The results from the commodity analysis reveal the fat-tailedness of the VAR innovations, since the estimated degrees of freedom of the t-distribution are very low. It turns out that the t-Lasso estimator, that accounts for this fat-tailedness, improves forecast accuracy. In our commodity analysis, we find volatility spillovers between energy and biofuels. Among energy commodities, most of the volatility spillovers involve gasoline, which is often blended with ethanol for consumption. Moreover, we observe volatility spillovers between energy and agriculture, regardless of the fact whether those crops can be used for biofuel production or not. We also find bidirectional volatility spillovers between biofuels and those agricultural commodities that can be used as inputs in biofuel production. These volatility spillovers are not observed in times of low energy prices, when investing in biofuels is less profitable.

The remainder of this article is structured as follows. Section 2 reviews the VAR model and introduces the corresponding t-Lasso estimator. Section 3 outlines the algorithm. The simulation study in Section 4 shows the good performance of the proposed estimator. Section 5 presents the data, the definition of volatility spillovers and the network analysis tool to visualize them. Section 6 presents the results of the volatility spillover analysis. Finally, Section 7 concludes.

2 Models and estimators

Let $\boldsymbol{y}_t$ be a J-dimensional vector of log volatilities for $1 \leq t \leq T$, with T the time series length. We take the logarithmic transformation of the volatilities, which ensures the positivity of the volatility forecasts (e.g. Callot et al., 2017). We consider a stationary VAR of order P for the log volatilities

$$\boldsymbol{y}_t = \sum_{p=1}^{P} \boldsymbol{B}_p \boldsymbol{y}_{t-p} + \boldsymbol{e}_t, \tag{1}$$

where $\boldsymbol{B}_p$ is the $J \times J$ matrix of autoregressive coefficients for lags $1 \leq p \leq P$, and $\boldsymbol{e}_t$ is a J-dimensional vector of error terms with zero mean and covariance matrix $\boldsymbol{\Sigma}$. Without loss of generality, we assume that all log volatility time series are mean centered such that no intercept is included.

For ease of notation, we rewrite model (1) in matrix form

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{B} + \boldsymbol{E},$$

where the $N \times J$ matrix $\boldsymbol{Y} = [\boldsymbol{y}'_{P+1}, \ldots, \boldsymbol{y}'_T]'$, with $N = T - P$. Let $\boldsymbol{X} = [\boldsymbol{X}_1, \ldots, \boldsymbol{X}_P]$ be the $N \times JP$ matrix of lagged time series, where $\boldsymbol{X}_p = [\boldsymbol{y}'_{P+1-p}, \ldots, \boldsymbol{y}'_{T-p}]'$ is a $N \times J$ matrix for $1 \leq p \leq P$. Finally, the $JP \times J$ matrix of autoregressive coefficients is $\boldsymbol{B} = [\boldsymbol{B}'_1, \ldots, \boldsymbol{B}'_P]'$ and $\boldsymbol{E}$ is the $N \times J$ error matrix.

In the next subsections we first review the penalized estimator for the Gaussian VAR model, and then we introduce the penalized estimator for the VAR model with Student t-errors. To this end, we build on the t-Lasso estimator for sparse graphical models proposed by Finegold and Drton (2011). We extend the approach of Finegold and Drton (2011) by (i) allowing for an unknown degrees of freedom of the t-distribution (ii) by adding a penalty on the inverse covariance matrix, ensuring existence of the estimator in any dimension (iii) extending their static framework for interdependencies between responses to the VAR model.

2.1 Gaussian innovations

Assume that the error terms $\boldsymbol{e}_t$ follow a multivariate normal distribution $N(\boldsymbol{0}, \boldsymbol{\Sigma})$. Consider the following penalized least squares estimator of $\boldsymbol{B}$

$$\widehat{\boldsymbol{B}} = \underset{\boldsymbol{B}}{\operatorname{argmin}} \frac{1}{2N} \operatorname{tr}\left[(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{B})(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{B})'\right] + \lambda \sum_{i,j=1}^{J} \sum_{p=1}^{P} |B_{p,ij}|, \tag{2}$$

where $\operatorname{tr}(\cdot)$ is the trace operator, $\lambda > 0$ is a regularization parameter associated to the l_1-penalty (Hastie et al., 2015, p8) on the ij^{th} entry of $\boldsymbol{B}_p$ denoted by $[\boldsymbol{B}_p]_{ij} = B_{p,ij}$. This penalty ensures that the estimation of $\boldsymbol{B}$ is feasible even if the number of parameters exceeds the time series length. Moreover, it sets some elements of $\widehat{\boldsymbol{B}}$ exactly equal to zero. The larger the regularization parameter λ, the sparser $\widehat{\boldsymbol{B}}$, that is the more of its elements are exactly zero.

In line with Rothman et al. (2010), we account for correlated errors by including the estimation of the inverse error covariance matrix $\boldsymbol{\Omega} = \boldsymbol{\Sigma}^{-1}$. To this end, we turn to the maximum likelihood framework and jointly estimate $\boldsymbol{B}$ and $\boldsymbol{\Omega}$ by minimizing the negative log likelihood:

$$(\widehat{\boldsymbol{B}}, \widehat{\boldsymbol{\Omega}}) = \underset{\boldsymbol{B},\boldsymbol{\Omega}}{\operatorname{argmin}} \frac{1}{2N} \operatorname{tr}\left[(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{B})\boldsymbol{\Omega}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{B})'\right] - \frac{1}{2}\log|\boldsymbol{\Omega}| + \lambda \sum_{i,j=1}^{J}\sum_{p=1}^{P}|B_{p,ij}| + \gamma \sum_{i\neq j}^{J}|\omega_{ij}|, \tag{3}$$

where ω_{ij} is the ij^{th} element of $\boldsymbol{\Omega}$, and $\gamma > 0$ is the regularization parameter associated to the l_1-penalty on the off-diagonal elements of the inverse error covariance matrix (Friedman et al., 2008). This l_1-penalty ensures that the estimation of $\boldsymbol{\Omega}$ is feasible even if the number of parameters exceeds the time series length. Furthermore, it sets some elements of $\widehat{\boldsymbol{\Omega}}$ to zero. The larger the regularization parameter γ, the sparser $\widehat{\boldsymbol{\Omega}}$. We refer to the estimator in (3) as the *Gaussian Lasso*.

2.2 Student *t*-innovations

We depart from the normality assumption and assume that the error terms are distributed according to a multivariate t-distribution $t_\nu(\mathbf{0}, \boldsymbol{\Psi})$, where $\nu > 0$ are the degrees of freedom and $\boldsymbol{\Psi}$ is the scale matrix, with associated variance-covariance matrix $\boldsymbol{\Sigma} = \boldsymbol{\Psi}\nu/(\nu - 2)$ if $\nu > 2$. The associated density function is given by

$$\frac{\Gamma((\nu + J)/2)|\boldsymbol{\Omega}|^{1/2}}{(\pi\nu)^{J/2}\Gamma(\nu/2)[1 + \boldsymbol{e}_t'\ \boldsymbol{\Omega}\ \boldsymbol{e}_t/\nu]^{(\nu+J)/2}}, \tag{4}$$

with $\boldsymbol{\Omega} = \boldsymbol{\Psi}^{-1}$ (Kotz and Nadarajah, 2004, p1). Recall that J is the dimension of the time series.

The Student t-distributed random vector $\boldsymbol{e}_t$ can be written as a Gaussian scale-mixture of $\boldsymbol{\phi}_t$ and τ_t (e.g. Kotz and Nadarajah, 2004, p2). Here $\boldsymbol{\phi}_t$ is a J-dimensional random vector distributed as a multivariate normal $N(\mathbf{0}, \boldsymbol{\Psi})$, independent of the random variable τ_t distributed as a Gamma $\Gamma(\nu/2, \nu/2)$. Then $\boldsymbol{e}_t = \boldsymbol{\phi}_t/\sqrt{\tau_t}$ follows a t-distribution $t_\nu(\mathbf{0}, \boldsymbol{\Psi})$. One can show that $\boldsymbol{e}_t$ has a normal conditional distribution

$$\boldsymbol{e}_t|\tau_t \sim N\left(\mathbf{0},\ \frac{\boldsymbol{\Psi}}{\tau_t}\right) \tag{5}$$

and

$$\tau_t|\boldsymbol{e}_t \sim \Gamma\left(\frac{\nu + J}{2},\ \frac{\nu + \boldsymbol{e}_t'\ \boldsymbol{\Omega}\ \boldsymbol{e}_t}{2}\right). \tag{6}$$

The arguments of the Gamma distribution are respectively the shape and scale parameters. By the properties of the Gamma distribution

$$E\left[\tau_t|\boldsymbol{e}_t\right] = \frac{\nu + J}{\nu + \boldsymbol{e}_t'\ \boldsymbol{\Omega}\ \boldsymbol{e}_t}. \tag{7}$$

The joint estimator of $\boldsymbol{B}$ and $\boldsymbol{\Omega}$ is defined as in (3), replacing the Gaussian by the t-density given in (4), and keeping the penalty terms. We call this estimator the *t-Lasso*.

3 Algorithm

We first assume that the degrees of freedom ν is known and discuss the EM (Expectation Maximization) algorithm to obtain the t-Lasso estimator, following Finegold and Drton (2011) for graphical models. However, in practice ν is not known and needs to be estimated. To this end, we use the ECM (Expectation Conditional Maximization) algorithm to include the estimation of ν (Liu and Rubin, 1995). The code of the algorithm is made available on the author's website.

3.1 EM algorithm with known ν

We treat τ_t as a hidden variable and want to estimate $\boldsymbol{B}$ and $\boldsymbol{\Omega}$. In the E-step, we compute the conditional expectation of the hidden variables according to (7). These expected values are put on the diagonal of an $N \times N$ diagonal matrix $\boldsymbol{\tau}$. In the M-step, we solve the optimization problem

$$(\widehat{\boldsymbol{B}}, \widehat{\boldsymbol{\Omega}} \,|\boldsymbol{\tau}) = \underset{\boldsymbol{B},\boldsymbol{\Omega}}{\operatorname{argmin}} \; \frac{1}{2N} \operatorname{tr}\left[\boldsymbol{\tau}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{B})\; \boldsymbol{\Omega}\; (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{B})'\right] - \frac{1}{2}\log|\boldsymbol{\Omega}| + \lambda \sum_{i,j=1}^{J} \sum_{p=1}^{P} |B_{p,ij}| + \gamma \sum_{i \neq j}^{J} |\omega_{ij}|, \quad (8)$$

where the conditional normality (5) of the error terms, given the hidden variables, is used. This M-step corresponds exactly to solving problem (3) for weighted observations. Algorithm 1 gives the details of the EM algorithm.

3.2 ECM algorithm with unknown ν

For unknown degrees of freedom ν, the M-step is replaced by two constrained maximizations (CM) steps, where first $(\boldsymbol{B}, \boldsymbol{\Omega})$ and then ν are estimated. An E-step is introduced before each CM-step, such that the weights are estimated twice in each iteration. This results is a multi-cycle version of the EM algorithm. Conditional on the latent variable τ_t for $t = 1, \ldots, N$, the log-likelihood of the parameters $\boldsymbol{B}$, $\boldsymbol{\Omega}$ and ν, ignoring constants, is:

$$L(\boldsymbol{B}, \boldsymbol{\Omega}, \nu|\boldsymbol{\tau}) = L_N(\boldsymbol{B}, \boldsymbol{\Omega}|\boldsymbol{\tau}) + L_G(\nu|\boldsymbol{\tau}), \quad (9)$$

where $L_N(\boldsymbol{B}, \boldsymbol{\Omega}|\boldsymbol{\tau})$ is the objective function in (8), and

$$L_G(\nu|\boldsymbol{\tau}) = -N \log\Gamma(\frac{\nu}{2}) + \frac{N\nu}{2}\log(\frac{\nu}{2}) + \frac{\nu}{2}\sum_{t=1}^{N}(\log(\tau_t) - \tau_t). \quad (10)$$

Given $\boldsymbol{B}$ and $\boldsymbol{\Omega}$, and assuming that the latent variable τ_t for $t = 1, \ldots, N$ is observed, we optimize (10) and obtain ν as the solution of

$$-\varphi(\frac{\nu}{2}) + \log(\frac{\nu}{2}) + \frac{1}{N}\sum_{t=1}^{N}(\log(\tau_t) - \tau_t) + 1 = 0, \quad (11)$$

Algorithm 1 t-Lasso with known ν: Expectation Maximization (EM)

Input $\mathbf{Y}$, $\mathbf{X}$, degrees of freedom ν and desired accuracy ε.

Initialization $\widehat{\boldsymbol{\Omega}}^{(0)} = \mathbf{I}_J$, $\widehat{\boldsymbol{B}}_p^{(0)} = \boldsymbol{I}_J$ for $1 \leq p \leq P$, and $\boldsymbol{e}_t^{(0)}$ the t^{th} row of $\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{B}}^{(0)}$.

Iteration Iterate the following steps for $m = 0, 1, 2, \ldots$:

E-step Compute for $t = 1, \ldots, N$ the weights

$$\widehat{\tau}_t^{(m+1)} = \frac{\nu + J}{\nu + (\boldsymbol{e}_t^{(m)})' \widehat{\boldsymbol{\Omega}}^{(m)} \boldsymbol{e}_t^{(m)}}.$$

M-step Compute $\widehat{\boldsymbol{B}}^{(m+1)}$ and $\widehat{\boldsymbol{\Omega}}^{(m+1)}$ using Algorithm A in Appendix A with inputs $\mathbf{Y} = \widehat{\boldsymbol{\tau}}^{*(m+1)}\boldsymbol{Y}$, and $\mathbf{X} = \widehat{\boldsymbol{\tau}}^{*(m+1)}\boldsymbol{X}$. Here, $\widehat{\boldsymbol{\tau}}^{*(m+1)}$ is the diagonal matrix having the square root of $\widehat{\tau}_t^{(m+1)}$ on its diagonal, for $t = 1, \ldots, N$.

Let $\boldsymbol{e}_t^{(m+1)}$ be the t^{th} row of $\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{B}}^{(m+1)}$.

Convergence Iterate until the relative change in the value of the objective function in (8) in two successive iterations is smaller than ε.

Output $\widehat{\boldsymbol{B}} = \widehat{\boldsymbol{B}}^{(m+1)}$ and $\widehat{\boldsymbol{\Omega}} = \widehat{\boldsymbol{\Omega}}^{(m+1)}$.

where $\varphi(\cdot)$ is the derivative of the log Gamma function. However, in practice τ_t is not observed. Thus, we replace the term $\sum_{t=1}^{N} (\log(\tau_t) - \tau_t)$ with its expectation (Liu and Rubin, 1995)

$$E[\sum_{t=1}^{N} (\log(\tau_t) - \tau_t) \,|\boldsymbol{B}, \boldsymbol{\Omega}, \nu] = \frac{1}{N}\sum_{t=1}^{N} (\log(\widehat{\tau}_t) - \widehat{\tau}_t) + \frac{1}{N}\left[\varphi\left(\frac{\nu + J}{2}\right) - \log\left(\frac{\nu + J}{2}\right)\right], \tag{12}$$

where we use (6) and where $\widehat{\tau}_t$ is computed as in (7). We substitute (12) in (11), and estimate ν as the solution of

$$-\varphi(\frac{\nu}{2}) + \log(\frac{\nu}{2}) + \frac{1}{N}\sum_{t=1}^{N} (\log(\widehat{\tau}_t) - \widehat{\tau}_t) + 1 + \frac{1}{N}\left[\varphi\left(\frac{\nu + J}{2}\right) - \log\left(\frac{\nu + J}{2}\right)\right] = 0. \tag{13}$$

Algorithm 2 presents the complete ECM algorithm.

4 Simulations

We analyze the performance of the *t-Lasso*, computed as outlined in Section 3. We use both the t-Lasso with fixed degrees of freedom (at the true value) and with estimated degrees of freedom. Their performance is compared to two alternative estimators: the *Least Squares* (LS), and the *Gaussian Lasso*, i.e. the solution of equation (3). The LS is the standard non-penalized estimator, the Gaussian Lasso is the benchmark for Gaussian models.

Algorithm 2 t-Lasso with unknown ν: Expectation Conditional Maximization (ECM)

Input $\mathbf{Y}$, $\mathbf{X}$, and desired accuracy ε.

Initialization $\widehat{\boldsymbol{\Omega}}^{(0)} = \mathbf{I}_J$, $\widehat{\boldsymbol{B}}_p^{(0)} = \boldsymbol{I}_J$ for $1 \leq p \leq P$, $\boldsymbol{e}_t^{(0)}$ the t^{th} row of $\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{B}}^{(0)}$, and $\widehat{\nu}^{(0)} = 1000$.

Iteration Iterate the following steps for $m = 0, 1, 2, \ldots$:

E-step 1 Compute for $t = 1, \ldots, N$ the weights:

$$\widehat{\tau}_t^{(m+\frac{1}{2})} = \frac{\widehat{\nu}^{(m)} + J}{\widehat{\nu}^{(m)} + (\boldsymbol{e}_t^{(m)})' \widehat{\boldsymbol{\Omega}}^{(m)} \boldsymbol{e}_t^{(m)}}.$$

CM-step 1 Compute $\widehat{\boldsymbol{B}}^{(m+1)}$ and $\widehat{\boldsymbol{\Omega}}^{(m+1)}$ using Algorithm A with inputs $\mathbf{Y} = \widehat{\boldsymbol{\tau}}^{*(m+\frac{1}{2})}\boldsymbol{Y}$, and $\mathbf{X} = \widehat{\boldsymbol{\tau}}^{*(m+\frac{1}{2})}\boldsymbol{X}$. Here, $\widehat{\boldsymbol{\tau}}^{*(m+\frac{1}{2})}$ is the diagonal matrix having the square root of $\widehat{\tau}_t^{(m+\frac{1}{2})}$ on its diagonal, for $t = 1, \ldots, N$.

Let $\boldsymbol{e}_t^{(m+1)}$ be the t^{th} row of $\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{B}}^{(m+1)}$.

E-step 2 Compute for $t = 1, \ldots, N$ the weights:

$$\widehat{\tau}_t^{(m+1)} = \frac{\widehat{\nu}^{(m)} + J}{\widehat{\nu}^{(m)} + (\boldsymbol{e}_t^{(m+1)})' \widehat{\boldsymbol{\Omega}}^{(m+1)} \boldsymbol{e}_t^{(m+1)}}.$$

CM-step 2 Compute $\widehat{\nu}^{(m+1)}$ with a one-dimensional search minimizing (13).

Convergence Iterate until the relative change in the value of the objective function in (8) in two successive iterations is smaller than ε.

Output $\widehat{\boldsymbol{B}} = \widehat{\boldsymbol{B}}^{(m+1)}$, $\widehat{\boldsymbol{\Omega}} = \widehat{\boldsymbol{\Omega}}^{(m+1)}$ and $\widehat{\nu} = \widehat{\nu}^{(m+1)}$.

Data generating process. We simulate from a VAR of order $P = 2$ with $J = 10$ time series. The dimensions of the VAR are in line with the ones in the application to be discussed in Section 5. The data generating process is:

$$\boldsymbol{y}_t = \boldsymbol{B}_1 \boldsymbol{y}_{t-1} + \boldsymbol{B}_2 \boldsymbol{y}_{t-2} + \boldsymbol{e}_t,$$

for $P + 1 \leq t \leq T = 100$. The autoregressive coefficient matrices are highly sparse in a structured manner: B_1 and B_2 have the same sparsity structure with non-zero elements (0.4 for B_1 and 0.2 for B_2) on the main diagonal and on the first row. The error terms $\boldsymbol{e}_t$ follow a multivariate Student $t_\nu(\mathbf{0}, \boldsymbol{\Psi})$, with $\nu \in \{1, 2, 3, 5, 10, \infty\}$. In the special case $\nu = 1$, the distribution is a multivariate Cauchy distribution, whereas for $\nu \to \infty$ the distribution is a multivariate normal. The ij^{th} entry of $\boldsymbol{\Psi}$ is $\psi_{i,j} = 0.1^{|i-j|}$, such that inverse error covariance matrix $\boldsymbol{\Omega}$ is a band matrix. We take $S = 1000$ simulations runs.

Table 1: Mean Absolute Estimation Error for the four estimators and different degrees of freedom ν.

ν	LS	Gaussian	t-Lasso	
		Lasso	ν fixed	ν estimated
1	0.638	0.187	0.088	0.089
2	0.135	0.096	0.088	0.088
3	0.107	0.090	0.088	0.089
5	0.098	0.090	0.088	0.089
10	0.095	0.091	0.089	0.090
∞	0.093	0.091	0.091	0.091

Performance measures. The different estimators are compared in terms of their estimation accuracy. To evaluate the estimation accuracy, we use the Mean Absolute Estimator Error

$$\text{MAEE}(\boldsymbol{B}, \widehat{\boldsymbol{B}}) = \frac{1}{S}\frac{1}{PJ^2}\sum_{S=1}^{S}\sum_{i,j=1}^{J}\sum_{p=1}^{P}|\widehat{B}_{p,ij}^{(s)} - B_{p,ij}^{(s)}|,$$

where $\widehat{B}_{p,ij}^{(s)}$ is the ij^{th} entry of the estimate $\widehat{\boldsymbol{B}}_p^{(s)}$ in simulation run s.

Simulation results. Table 1 reports the MAEEs for the four estimators and the different values of the true degrees of freedom ν. The t-Lasso with ν fixed at the true value always achieves the best MAEE. It is very closely followed by the t-Lasso with ν estimated. There is no considerable loss in estimation accuracy when estimating ν. Both t-Lasso estimators perform significantly better than the Gaussian Lasso (the difference in estimation accuracy is tested with a paired t-test, all p-values < 0.01). For instance, the improvement in estimation accuracy is of 53% and of 9% for $\nu = 1$ and $\nu = 2$ respectively. The margin by which the t-Lasso estimators outperform the Gaussian Lasso decreases for larger degrees of freedom. In particular, for $\nu = \infty$ there is no significant difference between the t-Lasso estimators and the Gaussian Lasso. Indeed, as $\nu \to \infty$, a multivariate t-distribution can be viewed as the approximation of a multivariate normal. The t-Lasso estimators also significantly outperform the LS for all values of ν: they improve estimation accuracy by, for instance, 18% for $\nu = 3$. The LS suffers from the large number of parameters to estimate, given the length of the time series.

Finally, Figure 1 shows the frequencies of the estimated degrees of freedom by the t-Lasso, for the different settings $\nu \in \{1, 2, 3, 10\}$. The estimated degrees of freedom are closely centered around the true value (vertical red line). The average (averaged over all simulation runs) estimated degrees of freedom are 0.94 (for $\nu = 1$), 2.10 (for $\nu = 2$), 3.16 (for $\nu = 3$) and 11.14 ($\nu = 10$). The variance of the estimated degrees of freedom increases for larger values of ν. We conclude that the degrees of freedom are quite accurately estimated.

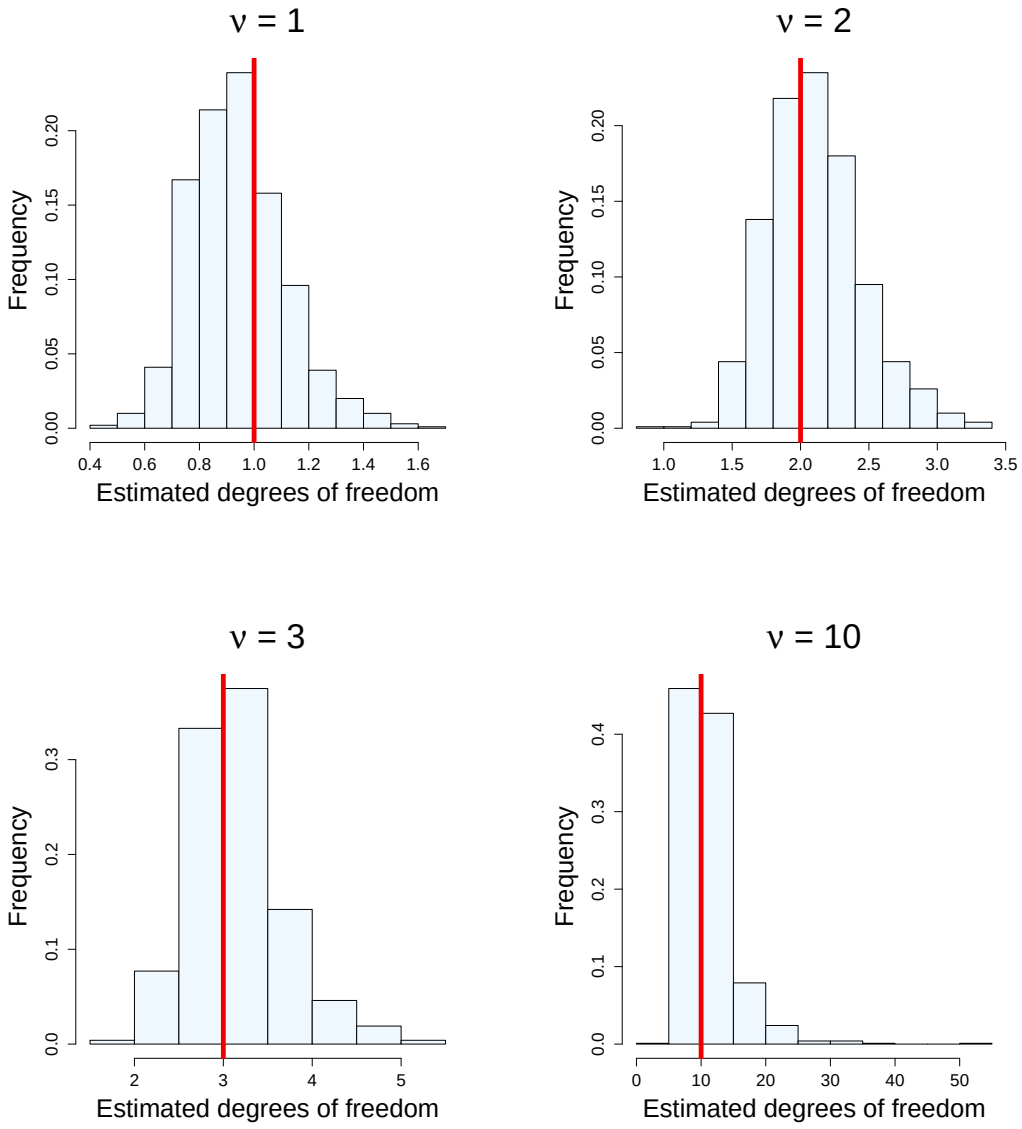

Figure 1: Frequency of the estimated degrees of freedom, for setting $\nu = 1, 2, 3, 10$ in the simulation study. The true value of the degrees of freedom is indicated by the vertical red line (dark gray on a gray scale).

Table 2: Variable description as in Thomson Reuters Eikon for Future Continuation 1 commodities.

Label	*Commodity*	*Description*
CRUO	Crude oil	NYMEX Light Sweet Crude Oil (WTI) Composite Energy
GASO	Gasoline	RBc1 NYMEX RBOB Gasoline Composite Energy
NATG	Natural gas	NGc1 NYMEX Henry Hub Natural Gas Composite Energy
ETHA	Ethanol	CBoT Denatured Fuel Ethanol Electronic Energy
CORN	Corn	Cc1 CBoT Corn Composite Commodity
WHEA	Wheat	Wc1 CBoT Wheat Composite Commodity
SOYB	Soybeans	Sc1 CBoT Soybeans Composite Commodity
SUGA	Sugar	ICE-US Sugar No. 11 Futures Electronic Commodity
COTT	Cotton	ICE-US Cotton No. 2 Futures Electronic Commodity
COFF	Coffee	KCc1 ICE-US Coffee C Futures Electronic Commodity

5 Data and volatility spillover analysis

In this section, we first present the data. Then, we define volatility spillovers based on the forecast error variance decomposition. Finally, we present a network analysis tool visualizing these volatility spillovers.

5.1 Data

We study $J = 10$ agricultural (corn, wheat, soybean, sugar, cotton, coffee), energy (crude oil, gasoline, natural gas) and biofuel (ethanol) commodities. We use daily information about the opening, highest and lowest prices of the nearest to maturity contracts traded in the corresponding future markets. Data are available on Thomson Reuters Eikon. The time span ranges from January 3^{rd} 2012 to October 28^{th} 2016, thus $T = 1070$ daily observations. Table 2 reports the label and the description of each variable. We obtain univariate volatility measures using the realized daily high-low range proposed by Parkinson (1980), see Appendix B for more details. We check for stationarity of the estimated volatilities with univariate unit root tests and the pooled unit root test of Levin et al. (2002) and find strong evidence in favor of stationarity (p-values < 0.01).

Second, we employ a rolling window over the period 2012-2016 with window size $W = 220$ days. At each time point $t = W, \ldots, T$, we estimate a VAR(P) model, with P selected using the Bayesian Information Criterion BIC, for the the log-volatilities from the date in the time window $t - W + 1$ until t. We assume that the error terms follow a multivariate t-distribution and estimate the degrees of freedom using the t-Lasso. Then, we study volatility spillovers based on forecast error variance decomposition, as described in the following subsection.

5.2 Spillover indices

We compute the volatility spillovers by means of the generalized forecast error variance decomposition, taking over the definitions from Diebold and Yilmaz (2015). Consider the Vector Moving Average (VMA) representation of the VAR(P) in model (1)

$$\boldsymbol{y}_t = \sum_{p=0}^{\infty} \boldsymbol{\theta}_p \boldsymbol{e}_{t-p},$$

where $\boldsymbol{\theta}_p$ is the moving average coefficient matrix at lag p (cfr. Wold's representation theorem, Lütkepohl, 2005, p25). Let $\widehat{\boldsymbol{y}}_{t+h}$ be the h-step ahead forecast for time $t+h$ made at time t with forecast horizon h. Then, the h-step-ahead forecast error for the j^{th} component of $\boldsymbol{y}_t$ is

$$\widehat{y}_{t+h,j} - y_{t+h,j} = \sum_{p=0}^{h-1} \theta_{p,j1} e_{t+h-p,1} + \ldots + \theta_{p,jJ} e_{t+h-p,J}, \tag{14}$$

where $e_{t,j}$ is the j^{th} component of $\boldsymbol{e}_t$, and $\theta_{p,jk}$ is the jk^{th} entry of $\boldsymbol{\theta}_p$.

If an impulse to $e_{t,k}$ of size of one standard deviation is given, then the expected value of the error term equals

$$E(\boldsymbol{e}_t | e_{t,k} = \sqrt{\sigma_{kk}}) = \frac{\boldsymbol{\Sigma}\boldsymbol{\delta}_k}{\sqrt{\sigma_{kk}}}, \tag{15}$$

where σ_{kk} is the kk^{th} entry of $\boldsymbol{\Sigma}$ and $\boldsymbol{\delta}_k$ is the selection vector of length J with unity entry as its k^{th} element and zeros elsewhere. Equation (15) defines a generalized impulse, as in Pesaran and Shin (1998), with response vector $\boldsymbol{\theta}_p \boldsymbol{\Sigma}\boldsymbol{\delta}_k / \sqrt{\sigma_{kk}}$, for $p = 1, \ldots, P$. It is important to note that (15) does not only hold for a normal distribution, but also for a t-distribution (e.g. Ding, 2016). If $\nu \leq 2$ and the covariance matrix of the t-distribution is not existing, the $\boldsymbol{\Sigma}$ in (15) should be replaced by the $\boldsymbol{\Psi}$ scale matrix.

The jk^{th} entry of the h-step ahead variance decomposition matrix is then defined as

$$w_{h,jk} = \frac{\sigma_{kk}^{-1} \sum_{p=0}^{h-1} (\boldsymbol{\delta}_j' \boldsymbol{\theta}_p \boldsymbol{\Sigma} \boldsymbol{\delta}_k)^2}{\sum_{p=0}^{h-1} \boldsymbol{\delta}_j' \boldsymbol{\theta}_p \boldsymbol{\Sigma} \boldsymbol{\theta}_p' \boldsymbol{\delta}_j}. \tag{16}$$

This is the proportion of the variance of the h-step-ahead forecast error (14) that is accounted for by the innovations in variable k of the VAR. Note that the denominator of (16) equals the variance of (14), and the numerator of (16) is the squared j^{th} component of the response vector. As the error term components are not orthogonal, in general $\sum_{k=1}^{J} w_{h,jk} \neq 1$. We consider the normalized variance decomposition

$$\tilde{w}_{h,jk} = \frac{w_{h,jk}}{\sum_{k=1}^{J} w_{h,jk}},$$

which is in the interval $[0, 1]$ by construction. The above normalized variance decomposition corresponds to the generalized variance decomposition proposed by Lanne and Nyberg (2016).

The *volatility spillover* from commodity k to commodity j is defined as

$$s_{h,k \to j} = 100 * \tilde{w}_{h,jk}. \tag{17}$$

The *volatility spillover index* is given by

$$s_h = \sum_{\substack{j \neq k \\ j,k=1}}^{J} s_{h,k \to j}, \tag{18}$$

as a single measure of the overall volatility spillovers and an overall proxy of the magnitude of the volatility spillovers. In the remainder, we take $h = 5$, which coincides to a working week (Bubák et al., 2011).

5.3 Network analysis

We visualize the volatility spillovers using a network analysis tool (e.g. Diebold and Yilmaz, 2015; Hautsch et al., 2015). The nodes in the network are the different commodities. An edge from commodity k to commodity j is drawn if $s_{h,k \to j}$ from equation (17) is non-zero; the edge width represents the magnitude of the volatility spillover. Hence, our network is directed and weighted, however, not necessarily sparse. Indeed, the sparsity of $\widehat{\boldsymbol{B}}$ is not necessarily preserved in the estimated VMA coefficient matrices $\widehat{\boldsymbol{\theta}}$. Consequently, the network will picture many non-zero volatility spillovers and hence contain many edges, although many of them may be quite small. Therefore, for better visualization, we give the network representing only the 15% largest volatility spillovers.

6 Results

In this section, we first present the time evolution of the estimated degrees-of-freedom and of the volatility spillovers (cfr. Section 5.1). Second, we picture the networks showing the volatility spillovers (cfr. Section 5.3). Finally, we show the good performance of the t-Lasso in terms of forecast accuracy.

6.1 Rolling window analysis

Estimated degrees of freedom. Figure 2 reports the estimated degrees of freedom of the multivariate t-distribution of the VAR residuals at each time point t, with t the end point of each time window. The average value of $\widehat{\nu}$ is 1.57, with maximum 1.80 and minimum 1.43. Overall, we observe that the estimated degrees of freedom are very low. This confirms the existence of heavy tails and justifies the use of the t-Lasso rather than the Gaussian Lasso.

If we look at the evolution of $\widehat{\nu}$ over time, we detect lower values of the estimated degrees of freedom in the time windows ending between the second half of 2014 and the first months of 2015. Volatile commodity markets characterized the second half of 2014, mainly driven by the drop of almost 50% in crude oil price (Knittel and Pindyck, 2016). Smaller values of $\widehat{\nu}$ indicate more pronounced extreme realizations in the VAR innovations and reflect a less predictable behavior of commodity markets.

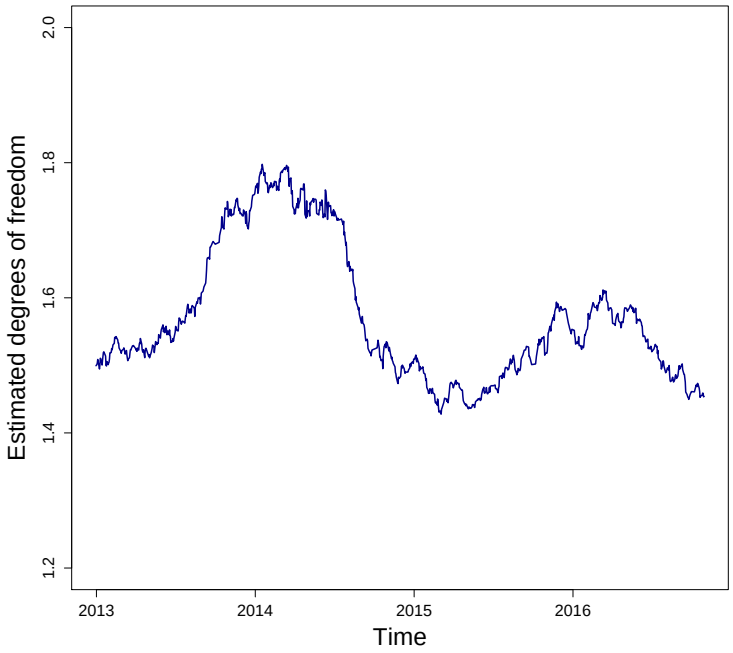

Figure 2: Estimated degrees of freedom of the multivariate t-distribution of the VAR residuals by the t-Lasso for a 220-day rolling window. The x-axis represents the ending date of each window.

Volatility spillover index. Figure 3 reports the evolution of the volatility spillover index (cfr. equation (18)) for the t-Lasso as a function of time t, being t the end point of each time window. We observe that the volatility spillover index experiences large swings over time. In particular, in the second half of 2014 we detect a large drop in the overall level of volatility spillovers. This drop is not permanent and during 2015 the volatility spillover index returns to the level prior to 2014.

The downturn in volatility spillover index can also be linked to the fall of crude oil price that occurred in the second half of 2014. Crude oil price experienced large downturns driven by (i) a larger supply from OPEC and non-OPEC countries and (ii) a weak global demand due to the slowdown of the world economy, notably the Chinese one. The drop in energy prices made biofuel less profitable and implied lesser and weaker volatility spillovers among energy, biofuel and agriculture commodities.

6.2 Network analysis

We visualize the volatility spillover in a network of commodities. Figure 4 top presents the network for the time window ending on October 28^{th} 2016, the last time point in our data set. For instance, a directed edge is drawn from gasoline (GASO) to crude oil (CRUO) since the volatility spillover from gasoline towards crude oil is different from zero and belongs to the largest 15% volatility spillovers. Recall that the edge width

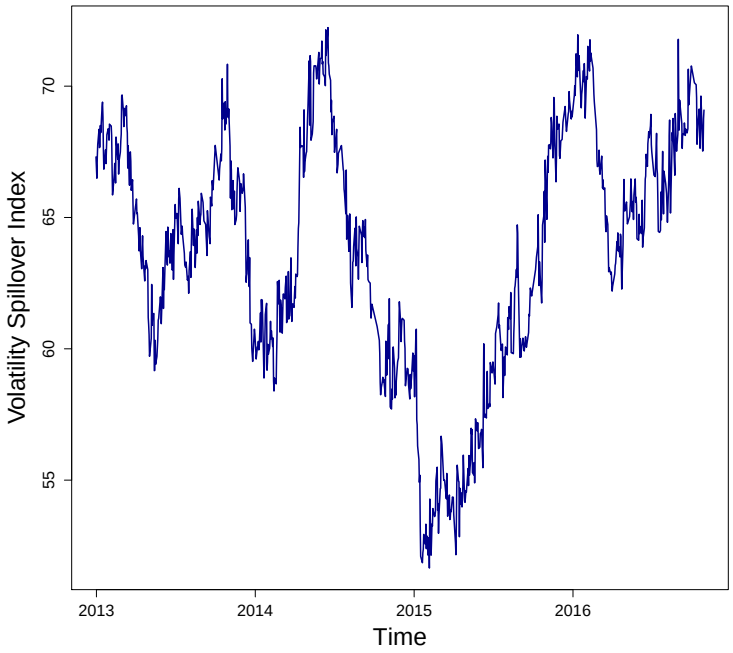

Figure 3: Volatility spillover index for the t-Lasso for a 220-day rolling window. The x-axis represents the ending date of each window.

represents the magnitude of the volatility spillover. Figure 4 also reports the networks for the time windows ending on February 5^{th} 2015 (bottom left) and on June 17^{th} 2014 (bottom right). For simplicity, from now on we refer to the three networks only by the year of the ending date of their windows. The 2015 and 2014 networks correspond to the time windows with the lowest and highest, respectively, values of the volatility spillover index.

Link energy-biofuel. In all networks we find volatility spillovers among energy and biofuel commodities. In the 2015 and 2014 networks, gasoline is the only energy commodity connected by an edge with ethanol. As ethanol is often blended in gasoline for consumption (Serra and Zilberman, 2013), it is not surprising to observe volatility spillovers between the two commodities in times of high energy prices (i.e. network 2014) and in times of large energy price changes (i.e. network 2015).

Link energy-agriculture. Bidirectional volatility spillovers between energy and agriculture are detected in all networks confirming the findings of Rezitis (2015). In the 2016 network, natural gas shows numerous and large spillovers from and towards agricultural commodities, whereas in the 2015 and 2014 networks, gasoline is the most connected energy commodity with agriculture. Both natural gas and gasoline are key commodities in the agricultural sector: the former is the major production cost for fertilizers, whereas the

latter is the most important fuel (together with diesel). If energy prices are high (e.g. network 2014), fuel has a large impact on agricultural commodities, which could explain the numerous volatility spillovers from/to gasoline. Conversely, as energy prices are low and fuel is cheap (e.g. network 2016), variations in fertilizer prices largely affect agricultural commodities volatilities, which could explain the many volatility spillovers from/to natural gas.

In all networks, various agricultural commodities are involved in volatility spillovers to and from energy. In general, volatility spillovers from and to energy involve both agricultural crops that are primary inputs for biofuel production - like wheat, soybean and sugar -, and crops that do not have a direct link with biofuels - like coffee -, as in Rezitis (2015).

Link biofuel-agriculture. Volatility spillovers between biofuel and agriculture are found only in the 2016 and 2014 networks. These volatility spillovers are bidirectional and involve wheat, sugar and cotton: this is consistent with our expectation since these crops can be used for ethanol production. In the 2015 network, we observe no volatility spillovers between biofuel and agriculture. In that period energy prices substantially dropped (Knittel and Pindyck, 2016): this made biofuel less attractive compare to other standard fuels and could have resulted in weaker volatility spillovers from and to agriculture.

6.3 Forecast accuracy

We now forecast the commodity volatilities. For a given forecast horizon, we forecast the vector of log volatilities $\widehat{\boldsymbol{y}}_{t+h}^{(t)} = [\widehat{y}_{t+h,1}^{(t)}, \ldots, \widehat{y}_{t+h,J}^{(t)}]$ based on the time window $[t-W+1, t]$, where t is the end point of the time window of size W and $\widehat{y}_{t+h,j}^{(t)}$ is the h-step ahead forecast of series j at time $t+h$ made at this end point. The forecast is obtained using the VAR model in (1), with autoregressive coefficients $\widehat{\boldsymbol{B}}$ obtained with the t-Lasso with ν estimated, the Gaussian Lasso and the LS. We compare these three estimators in terms of forecast accuracy by computing the Mean Absolute Forecast Error

$$\text{MAFE} = \frac{1}{N-h-W+1} \sum_{t=W}^{N-h} \text{MAFE}_t \quad \text{with} \quad \text{MAFE}_t = \frac{1}{J} \sum_{j=1}^{J} |\widehat{y}_{t+h,j}^{(t)} - y_{t+h,j}|.$$

Hence, the MAFE is computed for each time window and then averaged across all of them. The smaller the value of the mean absolute forecast error, the more accurate the volatility forecasts.

Results. Table 3 reports the results of the MAFE for the three estimators for different forecast horizons, namely $h = 1$, $h = 5$ and $h = 20$ days. The t-Lasso attains the best value of the MAFE for all forecast horizons. We find that based on the Diebold-Mariano test (Diebold and Mariano, 1995), the difference in forecast accuracy with the other estimators is always significant (p-values < 0.01). For forecast horizon $h = 1$, the t-Lasso gives a relative improvement in MAFE of 1% and of 23% over the Gaussian Lasso and

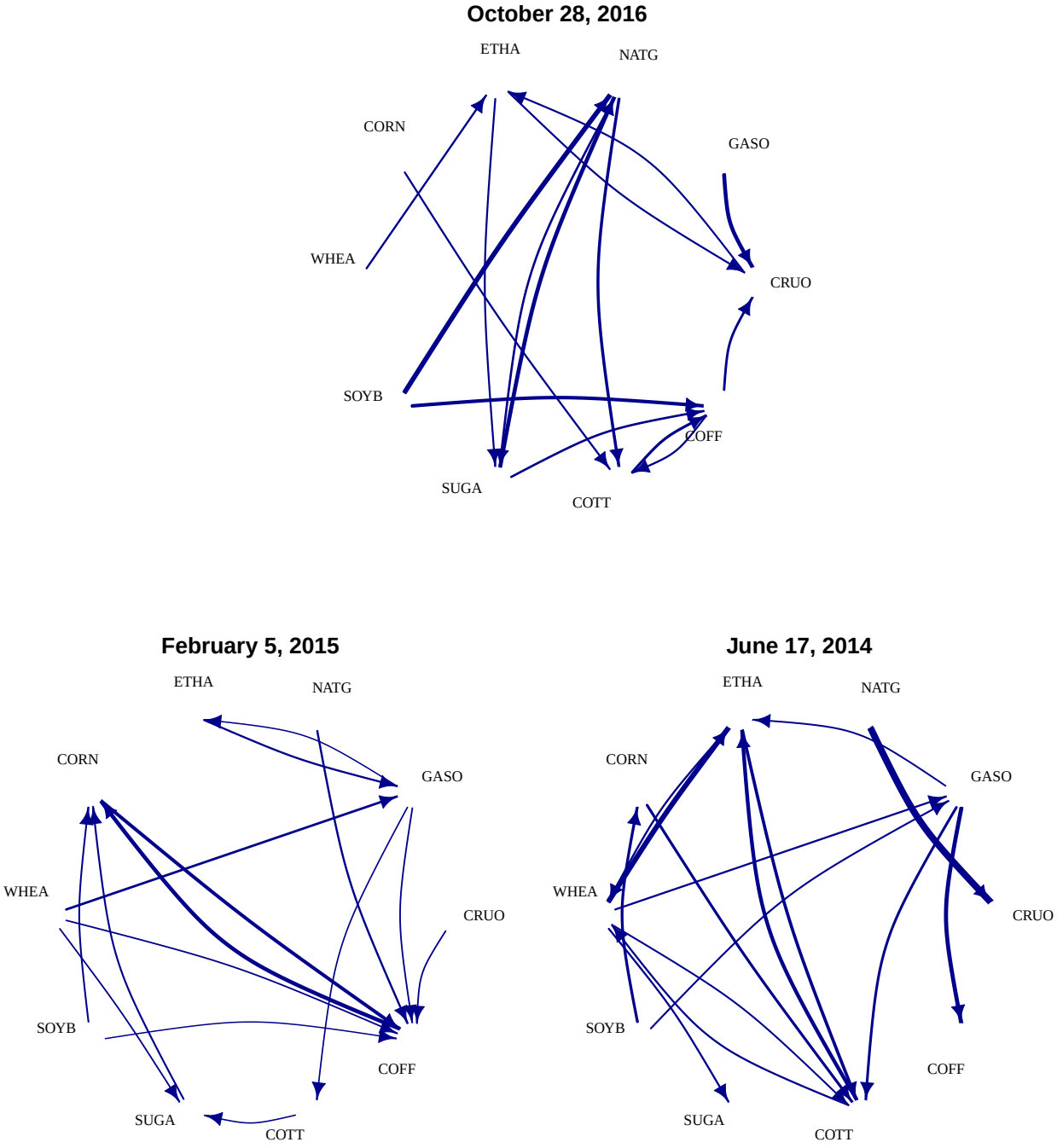

Figure 4: Commodity networks of volatility spillovers for different ending times of the 220-day rolling windows: October 28^{th} 2016 (top), February 5^{th} 2015 (bottom left) and June 17^{th} 2014 (bottom right).

Table 3: Mean Absolute Forecast Error for the t-Lasso, the Gaussian Lasso and the LS for forecast horizons $h = 1, 5$ and 20 from a 220-day rolling window.

Horizon	t-Lasso	Gaussian Lasso	LS
$h = 1$	1.871	1.896	2.418
$h = 5$	1.959	2.056	2.589
$h = 20$	2.280	2.636	2.820

the LS, respectively. The higher the forecast horizon, the better the performance of the t-Lasso compared to the Gaussian Lasso: for instance, for $h = 20$ our estimator gives an improvement of 14%.

Figure 5 reports the evolution of the MAFE_t (for $h = 20$) for the t-Lasso (blue solid line), the Gaussian Lasso (red dashed line) and the Least Squares (green dotted line) as a function of t, the end point of each time window. The t-Lasso attains a lower mean absolute forecast error than the Gaussian Lasso in all but five time windows and the difference in forecast accuracy is confirmed to be significant in all these time windows by the Diebold-Mariano test. Overall, we find that the t-Lasso attains a better forecast performance than the Gaussian Lasso and the LS, regardless of the forecast horizon or the presence of larger volatility spillovers.

7 Discussion

This paper studies volatility spillovers in a Vector AutoRegressive (VAR) model accounting for the fat-tailedness of its innovations. We extend the work of Diebold and Yilmaz (2015) and propose a penalized estimator of the VAR model with errors following a multivariate t-distribution, of which we include the estimation of the degrees of freedom. We emphasize that the normal distribution is a special case, for $\nu = \infty$. Our estimator is computable even for a large number of relatively short time series. Our simulation study shows that the proposed t-Lasso attains a better performance in terms of estimation accuracy than the Gaussian Lasso or Least Squares. The penalized estimator ensures good estimation accuracy even for VAR models that contain a large number of time series relative to the time series length. The t-distribution for the errors better captures the spikes typical of commodity volatility.

We study the dynamics of volatility spillovers between $J = 10$ energy, biofuel and agricultural commodities using a rolling window approach. Our findings highlight that the distributions of the log volatilities and of the residuals of the VAR model are fat-tailed. Furthermore, we visualize volatility spillovers using networks built on forecast error variance decomposition. We find evidence of bidirectional volatility spillovers between energy and agricultural commodities, regardless of the fact of being biofuel crops or not (Rezitis, 2015).

Our analysis is built on the volatility spillover definition by Diebold and Yilmaz (2014) and relies on

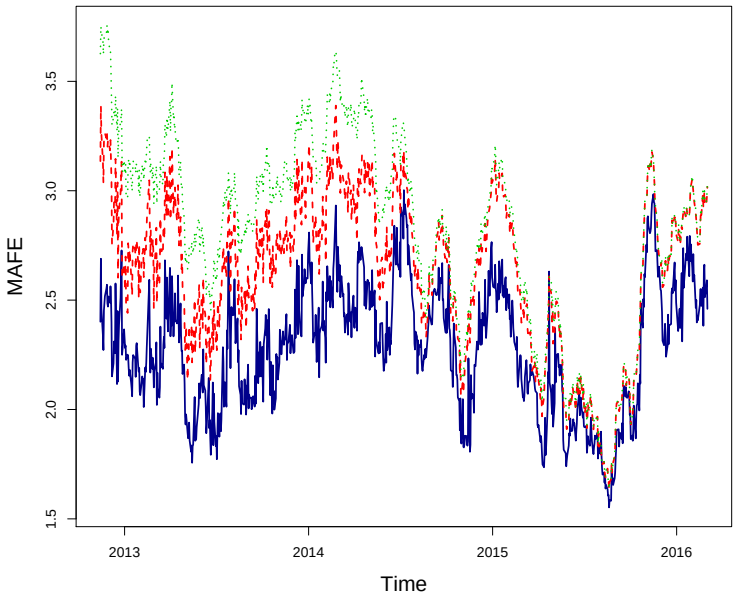

Figure 5: Mean Absolute Forecast Error for a 220-day rolling window and forecast horizon $h = 20$ for the t-Lasso (solid), the Gaussian Lasso (dashed) and the Leas Squares (dotted). The x-axis represents the ending date of each window.

generalized forecast error variance decomposition. Nevertheless, other definitions could have been used. For instance, we redid the analysis using Cholesky's decomposition (Diebold and Yilmaz, 2015, p14) and the spectral decomposition (Hafner and Herwartz, 2006). In both cases, we obtain results comparable to the ones reported, with only minor changes in the magnitudes of the volatility spillovers. The same holds when using other volatility spillover definitions. For instance, we redid the analysis defining a volatility spillover as the sum of squared generalized impulse responses, and draw comparable conclusions. Detailed results are available from the authors upon request.

Further research might follow several trajectories. On the one hand, one might investigate the existence of asymmetries in the VAR errors, for instance by considering a multivariate skewed t-distribution (Kotz and Nadarajah, 2004, p98) or by studying realized semi-variances as in Baruník et al. (2015). On the other hand, it could be interesting to use other data and volatility measures (McAleer and Medeiros, 2008). For commodity futures, data on the opening, lowest and highest daily prices are easier to obtain than intra-day high-frequency returns, motivating our choice of studying daily realized ranges. However, the innovation distribution could be less or more heavy-tailed depending on the asset type and/or the volatility measure, and this might have different implications in terms of risk management (for instance, see Brownlees and Gallo, 2010 for a Vale-at-Risk comparison between different realized measures of volatility for NYSE stocks).

Finally, the proposed t-VAR approach could be used in other research that study lagged effects among a large number of time series accounting for fat-tailed errors. Among other ones, the analysis of systemic risk, that is the risk that an individual firm default might threaten the stability of an entire market (Hautsch et al., 2015), seems a natural application. After the 2007-09, great importance has been given to the modeling of the spillovers among firms as a measure of systemic risk: typically, one would study a large number of firms (Hautsch et al., 2015), and model the non-normal innovation distribution caused by extreme downturns in the equity market (Engle et al., 2015). The proposed t-VAR approach would clearly address both issues, and its results would be highly interpretable in the form of risk networks.

Acknowledgments. This work was supported by the FWO (Research Foundation Flanders, contract number 12M8217N) and by the GOA/12/014 project of the Research Fund KU Leuven. We thank Piet Sercu and Kris Boudt for the useful comments that greatly improved the manuscript.

Appendix A: Algorithm Gaussian Lasso

The algorithm used to compute the Gaussian Lasso is given in Algorithm A . First, we solve for the autoregressive parameter $\boldsymbol{B}$ conditional on the inverse error covariance matrix $\boldsymbol{\Omega}$ using a coordinate descent algorithm, as in Friedman et al. (2007). Second we solve for $\boldsymbol{\Omega}$ conditional on $\boldsymbol{B}$, using the Graphical Lasso algorithm of Friedman et al. (2008). These two steps are implemented in `R` using the `grplasso` and `glasso` packages, respectively. We iteratively repeat the two steps until convergence of the objective function is reached.

Selection of regularization parameters. When solving for $\boldsymbol{B}|\boldsymbol{\Omega}$, we use a grid of regularization parameters λ and search for the optimal one minimizing the Bayesian Information Criterion (BIC)

$$\text{BIC}_\lambda = -2\text{log}L_\lambda + \text{df}_\lambda \text{log}(N),$$

where $\text{log}L_\lambda$ is the estimated likelihood, i.e. the first term in (3), using regularization parameter λ, and df_λ is the number of non-zero components of $\widehat{\boldsymbol{B}}_\lambda$. Likewise, when solving for $\boldsymbol{\Omega}|\boldsymbol{B}$ we use a grid of regularization parameters γ and search for the optimal one minimizing the BIC

$$\text{BIC}_\gamma = -2\text{log}L_\gamma + \text{df}_\gamma \text{log}(N),$$

where df_γ is the number of non-zero lower diagonal elements of $\widehat{\boldsymbol{\Omega}}$.

Algorithm A Gaussian Lasso: penalized estimation with Gaussian errors

Input $\mathbf{Y}$, $\mathbf{X}$, and desired accuracy ε.

Initialization $\widehat{\boldsymbol{\Omega}}^{(0)} = \mathbf{I}_J$.

Iteration Iterate the following steps for $m = 0, 1, 2, \ldots$:

Solving for $\boldsymbol{B}|\boldsymbol{\Omega}$. Compute $\widehat{\boldsymbol{B}}^{(m+1)}$:

$$\widehat{\boldsymbol{B}}^{(m+1)} = \underset{\boldsymbol{B}}{\operatorname{argmin}} \frac{1}{2N} \operatorname{tr}\left[(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{B})\widehat{\boldsymbol{\Omega}}^{(m)}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{B})'\right] + \lambda \sum_{i,j=1}^{J} \sum_{p=1}^{P} |B_{p,ij}|.$$

Solving for $\boldsymbol{\Omega}|\boldsymbol{B}$. Compute $\widehat{\boldsymbol{\Omega}}^{(m+1)}$:

$$\widehat{\boldsymbol{\Omega}}^{(m+1)} = \underset{\boldsymbol{\Omega}}{\operatorname{argmin}} \frac{1}{2N} \operatorname{tr}\left[(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{B}}^{(m+1)})\boldsymbol{\Omega}(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{B}}^{(m+1)})'\right] - \frac{1}{2}\log|\boldsymbol{\Omega}| + \gamma \sum_{i \neq j}^{J} |\omega_{ij}|.$$

Convergence Iterate until the relative change in the value of the objective function in (3) in two successive iterations is smaller than ε.

Output $\widehat{\boldsymbol{B}} = \widehat{\boldsymbol{B}}^{(m+1)}$ and $\widehat{\boldsymbol{\Omega}} = \widehat{\boldsymbol{\Omega}}^{(m+1)}$.

Appendix B: Volatility measure

Following Parkinson (1980), we obtain a measure of volatility of a future contract using the high-low daily range estimator. Consider the daily information about opening price $O_{t,j}$, the highest price $H_{t,j}$ and the lowest price $L_{t,j}$ attained on date $1 \leq t \leq T$ for commodity $1 \leq j \leq J$. The high-low range estimator for the daily variance is

$$v_{t,j} = \frac{1}{4\log(2)}(h_{t,j} - l_{t,j})^2,$$

where $h_{t,j} = \log(H_{t,j}) - \log(O_{t,j})$ and $l_{t,j} = \log(L_{t,j}) - \log(O_{t,j})$ are the maximum and minimum daily return, respectively. For a review on realized range measures of volatility see Shu and Zhang (2006) or Martens and van Dick (2007). The time series entering the VAR of equation (1) are the log trasformations of the realized ranges $\widehat{v}_{t,j}$ of commodity j at time t, that is $\mathbf{y}_t = [\log(\widehat{v}_{t,j}), \ldots, \log(\widehat{v}_{t,J})]'$. The original volatility series can be obtained by applying the exponent transformation and correction for the re-transformation bias as in Bauer and Vorkink (2011).

References

Andersen, T. G.; Bollerslev, T.; Diebold, F. X. and Heiko, E. (2001), "The distribution of realized stock return volatility," *Journal of Financial Economics*, 61(1), 43–76.

Barigozzi, M. and Hallin, M. (2017), "A network analysis of the volatility of high dimensional financial series," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 66(3), 581–605.

Barndorff-Nielsen, O. E. and Shephard, N. (2002), "Econometric analysis of realized volatility and its use in estimating stochastic volatility models," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(2), 253–280.

Baruník, J.; Kočenda, E. and Vácha, L. (2015), "Volatility spillovers across petroleum markets," *The Energy Journal*, 36(3), 309–330.

Basu, S. and Michailidis, G. (2015), "Regularized estimation in sparse high-dimensional time series models," *The Annals of Statistics*, 43(4), 1535–1567.

Bauer, G. H. and Vorkink, K. (2011), "Forecasting multivariate realized stock market volatility," *Journal of Econometrics*, 160(1), 93–101.

Brownlees, C. T. and Gallo, G. M. (2010), "Comparison of volatility measures: A risk management perspective," *Journal of Financial Econometrics*, 8(1), 29–56.

Bubák, V.; Kočenda, E. and Žikeš, F. (2011), "Volatility transmission in emerging European foreign exchange markets," *Journal of Banking and Finance*, 35(11), 2829–2841.

Callot, L. A. F.; Kock, A. B. and Medeiros, M. C. (2017), "Modeling and forecasting large realized covariance matrices and portfolio choice," *Journal of Applied Econometrics*, 32(1), 140–158.

Caporin, M. and Velo, G. G. (2015), "Realized range volatility forecasting: Dynamic features and predictive variables," *International Review of Economics and Finance*, 40, 98–112.

Christensen, K.; Podolskij, M. and Vetter, M. (2009), "Bias-correcting the realized range-based variance in the presence of market microstructure noise," *Finance and Stochastics*, 13(2), 239–268.

Corsi, F.; Mittnik, S.; Pigorsch, C. and Pigorsch, U. (2008), "The volatility of realized volatility," *Econometric Reviews*, 27(1-3), 46–78.

Davis, R.; Zang, P. and Zheng, T. (2016), "Sparse vector autoregressive modeling," *Journal of Computational and Graphical Statistics*, 25(4), 1077–1096.

Derimer, M.; Diebold, F. X.; Liu, L. and Yilmaz, K. (2017), "Estimating Global Bank Network Connectedness," *Forthcoming in Journal of Applied Econometrics*.

Diebold, F. X. and Mariano, R. S. (1995), "Comparing predictive accuracy," *Journal of Business and Economic Statistics*, 13(3), 253–263.

Diebold, F. X. and Yilmaz, K. (2014), "On the network topology of variance decompositions: Measuring the connectedness of financial firms," *Journal of Econometrics*, 182(1), 119–134.

— (2015), *Financial and macroeconomics connectedness: A network approach to measurement and monitoring*, Oxford University Press, New York, US.

Ding, P. (2016), "On the conditional distribution of the multivariate t distribution," *The American Statistician*, 70(3), 293–295.

Engle, R.; Jondeua, E. and Rockinger, M. (2015), "Systemic risk in Europe," *Review of Finance*, 19 (1), 145–190.

Finegold, M. and Drton, M. (2011), "Robust graphical modeling of gene networks using classical and alternative t-distributions," *The Annals of Applied Statistics*, 5(2A), 1057–1080.

Franses, P. and Lucas, A. (1998), "Outlier detection in cointegration analysis," *Journal of Business & Economic Statistics*, 16(4), 459–468.

Friedman, J.; Hastie, T.; Höfling, H. and Tibshirani, R. (2007), "Pathwise Coordinate Optimization," *The Annals of Applied Statistics*, 1(2), 302–332.

Friedman, J.; Hastie, T. and Tibshirani, R. (2008), "Sparse inverse covariance estimation with the graphical lasso," *Biostatistics*, 9(3), 432–441.

Gelper, S.; Wilms, I. and Croux, C. (2016), "Identifying demand effects in a large network of product categories," *Journal of Retailing*, 92(1), 25–39.

Hafner, C. M. and Herwartz, H. (2006), "Volatility impulse responses for multivariate GARCH models: An exchange rate illustration," *Journal of International Money and Finance*, 25(5), 719–740.

Hassler, U.; Rodrigues, P. M. M. and Rubia, A. (2016), "Quantile regression for long memory testing: A case of realized volatility," *Journal of Financial Econometrics*, 14(4), 693–724.

Hastie, T.; Tibshirani, R. and Wainwright, M. (2015), *Statistical learning with sparsity: The lasso and generalizations*, CRC press.

Hautsch, N.; Schaumburg, J. and Schienle, M. (2015), "Financial network systemic risk contributions," *Review of Finance*, 19(2), 685–738.

Knittel, C. R. and Pindyck, R. S. (2016), "The simple economics of commodity price speculation," *American Economic Journal: Macroeconomics*, 8(2), 85–110.

Kotz, S. and Nadarajah, S. (2004), *Multivariate t distributions and their applications*, Cambridge University Press, Cambridge, UK.

Lanne, M. and Nyberg, H. (2016), "Generalized forecast error variance decomposition for linear and nonlinear multivariate models," *Oxford Bulletin of Economics and Statistics*, 78(4), 595–603.

Levin, A.; Lin, C.-F. and Chu, C.-S. J. (2002), "Unit root tests in panel data: Asymptotic and finite-sample properties," *Journal of Econometrics*, 108(1), 1–24.

Liu, C. and Rubin, D. B. (1995), "ML estimation of the t distribution using EM and its extensions, ECM and ECME," *Statistica Sinica*, 5, 19–39.

Lütkepohl, H. (2005), *New introduction to multiple time series analysis*, Springer, Heidelberg, Germany.

Martens, M. and van Dick, D. (2007), "Measuring volatility with the realized range," *Journal of Econometrics*, 138(1), 181–207.

McAleer, M. and Medeiros, M. C. (2008), "Realized volatility: A review," *Econometric Reviews*, 27(1-3), 10–45.

Parkinson, M. (1980), "The extreme value method for estimating the variance of the rate of return," *The Journal of Business*, 53(1), 61–65.

Pesaran, H. H. and Shin, Y. (1998), "Generalized impulse response analysis in linear multivariate models," *Economics Letters*, 58(1), 17–29.

Rezitis, A. N. (2015), "The relationship between agricultural commodity prices, crude oil prices and US dollar exchange rates: A panel VAR approach and causality analysis," *International Review of Applied Economics*, 29(3), 403–434.

Rothman, A. J.; Levina, E. and Zhu, J. (2010), "Sparse multivariate regression with covariance estimation," *Journal of Computational and Graphical Statistics*, 19 (4), 947–962.

Serra, T. (2011), "Volatility spillovers between food and energy markets: A semiparametric approach," *Energy Economics*, 33, 1155–1164.

Serra, T. and Zilberman, D. (2013), "Biofuel-related price transmission literature: A review," *Energy Economics*, 37, 141–151.

Shu, J. and Zhang, J. E. (2006), "Testing range estimators of historical volatility," *The Journal of Futures Markets*, 26(3), 297–313.

Tibshirani, R. (1996), "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 58(1), 267–288.

FACULTY OF ECONOMICS AND BUSINESS
Naamsestraat 69 bus 3500
3000 LEUVEN, BELGIË
tel. + 32 16 32 66 12
fax + 32 16 32 67 91
info@econ.kuleuven.be
www.econ.kuleuven.be